Conference Abstract

Towards a Global Collection Description Standard

Niels Raes[‡], Emily van Egmond[§], Ana B. Casino[|], Matt Woodburn[¶], Deborah L Paul[#]

‡ NLBIF, Leiden, Netherlands
§ Free University, Amsterdam, Netherlands
| CETAF c/o Royal Belgian Institute of Natural Sciences, Brussels, Belgium
¶ Natural History Museum, London, United Kingdom
# Florida State University, Tallahassee, United States of America

Abstract

With digitisation of natural history collections over the past decades, their traditional roles — for taxonomic studies and public education — have been greatly expanded into the fields of biodiversity assessments, climate change impact studies, trait analyses, sequencing, 3D object analyses etc. (Nelson and Ellis 2019; Watanabe 2019). Initial estimates of the global natural history collection range between 1.2 and 2.1 billion specimens (Ariño 2010), of which 169 million (8-14% - as of April 2019) are available at some level of digitisation through the Global Biodiversity Information Facility (GBIF). With iDigBio (Integrated Digitized Biocollections) established in the United States and with the European DiSSCo (Distributed Systems of Scientific Collections) accepted on the ESFRI roadmap, it has become a priority to digitize natural history collections at an industrialized scale. Both iDigBio and DiSSCo aim at mobilising, unifying and delivering bio- and geo-diversity information at the scale, form and precision required by scientific communities, and thereby transform a fragmented landscape into a coherent and responsive research infrastructure. In order to prioritise digitisation based on scientific demand, and efficiency using industrial digitisation pipelines, it is required to arrive at a uniform and unambiguously accepted collection description standard that would allow comparing, grouping and analysing natural history collections at diverse levels.

Several initiatives attempt to unambiguously describe natural history collections using taxonomic and storage classification schemes. These initiatives include One World

Collection, Global Registry of Scientific Collections (GRSciColl), TDWG (Taxonomic Databases Working Group) Natural Collection Descriptions (NCD) and CETAF (Consortium of European Taxonomy Facilities) passports, among others. In a collaborative effort of DiSSCo, ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage), iDigBio, TDWG and the Task Group Collection Digitisation Dashboards, the various schemes were compared in a cross-walk analysis to propose a preliminary natural collection description standard that is supported by the wider community. In the process, two main user groups of collection descriptions standards were identified; scientists and collection managers. The classification produced intends to meet requirements from them both, resulting in three classification schemes that exist in parallel to each other (van Egmond et al. 2019). For scientific purposes a 'Taxonomic' and 'Stratigraphic' classification were defined, and for management purposes a 'Storage' classification. The latter is derived from specimen preservation types (e.g. dried, liquid preserved) defining storage requirements and the physical location of specimens in collection holding facilities. The three parallel collection classifications can be cross-sectioned with a 'Geographic' classification to assign sub-collections to major terrestrial and marine regions, which allow scientists to identify particular taxonomic or stratigraphic (sub-)collections from major geographical or marine regions of interest.

Finally, to measure the level of digitisation of institutional collections and progress of digitisation through time, the number of digitised specimens for each geographically cross-sectioned (sub-)collection can be derived from institutional collection management systems (CMS). As digitisation has different levels of completeness a 'Digitisation' scheme has been adopted to quantify the level of digitisation of a collection from Saarenmaa et al. 2019, ranging from 'not digitised' to extensively digitised, recorded in a progressive scale of MIDS (Minimal Information for Digital Specimen).

The applicability of this preliminary classification will be discussed and visualized in a Collection Digitisation Dashboards (CDD) to demonstrate how the implementation of a collection description standard allows the identification of existing gaps in taxonomic and geographic coverage and levels of digitisation of natural history collections. This set of common classification schemes and dashboard design (van Egmond et al. 2019) will be contributed to the TDWG Collection Description interest group to ultimately arrive at the common goal of a 'World Collection Catalogue'.

Keywords

TDWG CD, TDWG, CETAF, DiSSCo, iDigBio, ICEDIG, collection standards

Presenting author

Niels Raes

Presented at

Biodiversity_Next 2019

References

- Ariño A (2010) Approaches to estimating the universe of natural history collections data. Biodiversity Informatics 7 (2). https://doi.org/10.17161/bi.v7i2.3991
- Nelson G, Ellis S (2019) The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B: Biological Sciences 374 (1763). https://doi.org/10.1098/rstb.2017.0391
- Saarenmaa H, Agosti D, Dillen M, Egloff W, Gagnier P, Groom Q, Hardisty A, Raes N (2019) Report on implementation of open access policies in collection institutions. MILESTONE – MS35. Zenodo https://doi.org/10.5281/ZENODO.2632739
- van Egmond E, Willemse L, Paul D, Woodburn M, Casino A, Gödderz K, Vermeersch X, Bloothoofd J, Wijers A, Raes N (2019) Design of a Collection Digitisation Dashboard. Zenodo https://doi.org/10.5281/ZENODO.2621055
- Watanabe ME (2019) The Evolution of Natural History Collections. BioScience 69 (3): 163-169. https://doi.org/10.1093/biosci/biy163